A sentiment analysis dataset is provided to you.

It has a training dataset and a test data set

Task1:

1. Train a model on training data set to predict a positive or negative review (1 for positive, 0 for negative)
2. You can use BOW or TFIDF vectorizing technique to do feature extraction
3. Evaluate your score (classification report from Sklearn metrics) on the test data set
4. Before training, apply the following cleaning steps:
   1. Remove mentions i.e. any alphanumeric starting with “@”
   2. Remove “n\”
   3. Remove all links i.e. any alphanumeric starting with https or http
   4. Remove all hashtags i.e. “#”

Task2:

Using the cleaned data:

1. Use text blob to predict (on test) a positive or negative review & generate classification report
2. Use VADER to predict (on test) a positive or negative review & generate classification report
3. Use Hugging Face to predict (on test) a positive or negative review & generate classification report
4. Use your trained model to predict (on test) a positive or negative review & generate classification report
5. Rank the results, how did your trained model did w.r.t to all others?

**Note: for simplicity, consider neutral review as positive i.e. if VADER score is zero (which is neutral), consider it positive**